DOCUMENT RESUME

ED 362 552 TM 020 586

AUTHOR Westers, Paul; Kelderman, Henk
TITLE Generalizations of the Solution-Error Response-Error Model. Project Psychometric Aspects of Item Banking No. 54. Research Report 93-1.
INSTITUTION Twente Univ., Enschede (Netherlands). Dept. of Education.
PUB DATE Apr 93
NOTE 45p.
AVAILABLE FROM Bibliotheek, Department of Education, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Equations (Mathematics); *Estimation (Mathematics); Foreign Countries; Generalization; *Goodness of Fit; *Item Response Theory; *Mathematical Models; Pathology
IDENTIFIERS *Latent Class Models; Multidimensionality (Tests); Polytomous Scoring; *Solution Error Response Error Model

ABSTRACT

In the last decade, several attempts have been made to relate item response theory (IRT) models to latent class analysis (LCA) models. One of these attempts is the solution-error response-error (SERS) model, an LCA model in which the structure of the latent class probabilities is explained by a one-dimensional loglinear Rasch model. In this paper, the SERE model is generalized to models for polytomously scored latent states that may be explained by a multidimensional latent space. In this generalized SERE model a distinction is made between some well-defined latent states in which the subject has a certain amount of knowledge of the answer. The probability that the subject is in a certain state is assumed to be governed by the multidimensional polytomous latent trait model. The relationship between the latent states and the observed answers is described by conditional probabilities. Two appendixes present items from a clinical pathology test and a proof of the collapsed generalized SERE model. Five tables and four figures illustrate the discussion. (Contains 38 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 362 552

Generalizations of the Solution-Error Response-Error Model

Research Report 93-1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. NELISSEN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paul Westers

Henk Kelderman

TM020586

department of EDUCATION

University of Twente

Division of Educational Measurement and Data Analysis

BEST COPY AVAILABLE

Project Psychometric Aspects of Item Banking No. 54

Colofon:
Typing: P. Westers and L.A.M. Bosch-Padberg
Cover design: Audiovisuele Sectie TOLAB
Toegepaste Onderwijskunde
Printed by: Centrale Reproductie-afdeling
Oplage: 120

Generalizations of the Solution-Error Response-Error Model

Paul Westers

Henk Kelderman

The authors thank Mary E.Lunz and Barbara M. Castleberry of the American Society of Clinical Pathologist (ASCP) for providing empirical data used in the example in this paper.

The address of Paul Westers is Centre of Biostatistics, Department of Biology, University of Utrecht, Padualaan 14, 3584 CH Utrecht, The Netherlands.

Generalizations of the solution-error response-error model , Paul Westers and Henk Kelderman - Enschede : University of Twente, Department of Education, April, 1993. - 38 pages

Abstract

In the last decade, several attempts have been made to relate item response theory (IRT) models to latent class analysis (LCA) models. One of these attempts is the solution-error response-error (SERE) model; a LCA model in which the structure of the latent class probabilities is explained by a one-dimensional loglinear Rasch model.

In this paper the SERE model will be generalized to models for polytomously scored latent states that may be explained by a multidimensional latent space.

Generalizations of the Solution-Error Response-Error Model

Introduction

For measuring individual differences, a distinction can be made between measurements on a discrete qualitative latent trait and measurements on a continuous quantitative scale. The latent class analysis (LCA) model, where it is assumed that subjects belong to different latent classes, belongs to the first case (Bartholomew, 1987; Lazarsfeld & Henry, 1968; Mooijaart, 1978). Whereas, the item response models (IRT) belongs to the second case. Some well known examples of IRT models are the Rasch (1960/1980) model and the two and three-parameter-logistic or normal ogive models (Lord, 1980; Lord & Novick, 1968). In the last decade, several attempts have been made to relate IRT models to LCA models (Bock & Aitkin, 1981; Dayton & Macready, 1980; Formann, 1985; Kelderman, 1988a, 1989; Kelderman & Macready, 1990; Mislevy & Verhelst, 1990; Yamamoto, 1987, 1988). In this paper one of these attempts will be discussed: the solution-error response-error (SERE) model of Kelderman (1988a).

In the SERE model a distinction is made between a "Know" state in which the subject has complete knowledge of the answer and a "Don't know" state. The probability that the subject is in the "Know" state is assumed to be governed by the Rasch model. Further, it is assumed that if the subject is in the "Don't know" state, the subject will choose an alternative, where the attractiveness may be different for different alternatives, including the correct one. The SERE model can be formulated as an (incomplete) LCA model, where each latent class corresponds to an idealized response pattern. The relations between these idealized responses are explained by the loglinear version of the Rasch model (Cressie & Holland, 1983; Duncan, 1984; Kelderman, 1984; Tjur, 1982).

All SERE models considered in Kelderman (1988a) deal with a one-dimensional continuous latent trait. In many testing situations, however, we may have to deal with a two - or more - dimensional latent space. For example, consider a version of the American Society of Clinical Pathologist (ASCP) Microbiology-Test. In Appendix A.1 some items of this test are presented. Content experts have hypothesized that although each item of this ASCP test has one correct alternative, incorrect responses might often be chosen after cognitive activities similar to those necessary to arrive at the correct response. They presumed further that "Applying Knowledge", "Selecting Action", "Calculating", "Correlating Data" and "Evaluating Problem" are the involved cognitive processes in answering the items. For instance, it was assumed that for item 11 of Appendix A.1 the correct answer (d) involved two applications of knowledge, whereas answer c only one. For giving the correct answer c on item 20 it was assumed that the subject had to use the cognitive process "Evaluating Problem" two times and the cognitive processes "Applying Knowledge" and "Selecting Action" ones.

So, in general, producing one answer may require quite another ability from the examinee than to produce another answer. Or some responses may require the repeated application of an ability, whereas others may require only a single application of the same ability. In this paper the SERE model will be generalized to models for polytomously scored latent states that may be explained by a multidimensional latent space. Maximum likelihood estimates of the parameters of this generalized SERE (GSERE) model can be obtained by solving the likelihood equations by the iterative proportional fitting (IPF) algorithm of Goodman (1974b).

In what follows, the GSERE model will be formulated. The estimation method and goodness-of-fit tests are described, and the question of identifiability is discussed.

The Generalized Solution-Error Response-Error Model

Suppose that each subject, randomly drawn from a population of subjects, responds to k test items, where the answer to item j may be any of the r_j responses, denoted by y_j (y_j=1,...,r_j). Let x_j (x_j=0,...,s_j) indicate the latent state of the subject. For example, all the items of the ASCP Microbiology Test have four possible responses (i.e., r_j=4) and may have three latent states: "Don't know", "Partial knowledge" and "Complete knowledge". It is assumed that if the subject is in the "Don't know" state (x_j=0), (s)he will choose one of the alternatives. If the subject is in the "Partial knowledge" state (x_j=1), (s)he will choose one of the alternatives that might be right given the subject's partial knowledge of the answer. If the subject is in the "Complete knowledge" state (x_j=2), (s)he will choose the right alternative. The random variables with values y_j and x_j are denoted by Y_j and X_j (j=1,...,k). The relationship between the latent state x_j and the observed response y_j is described by the conditional probability

$$(1) \qquad \Phi^{X_j Y_j}_{x_j y_j} \equiv P(Y_j=y_j|X_j=x_j) ,$$

This conditional probability will be referred to as the attraction parameter of item j.

In the generalized solution-error response-error (GSERE) model it is assumed that the latent states are not governed by the Rasch (1960/1980) model, but by the more general multidimensional polytomous latent trait model (MPLT) by Kelderman (1988b). In the MPLT model it is assumed that the subject must perform certain cognitive operations to produce a score x_j on item j. See for instance the example in the previous section. Each operation depends on a certain proficiency on a latent trait. Let $B_{jq}(x)$ be a non-negative weight associated with the dependence of response x on item j on the latent trait q. Furthermore, let $\delta_{jq}(x)$ be the difficulty parameter of the response x on item j related to latent trait q (q=1,...,v), θ_q be a value of the

subject on the latent ability continuum and $\theta = (\theta_1,....,\theta_v)$ be the vector of ability values. The probability that the subject has a response x_j on item j can be written as (Kelderman, 1988b)

$$(2) \qquad P(x_j|\theta) = \frac{\exp\{\sum_q (\theta_q - \delta_{jq}(x_j))B_{jq}(x_j)\}}{\sum_z \exp\{\sum_q (\theta_q - \delta_{jq}(z))B_{jq}(z)\}}$$

Assuming local independence of X_j and Y_j given the latent trait vector θ and X_j, respectively, the probability of choosing response y_j is equal to

$$(3) \qquad P(y_j|\theta) = \sum_{x_j} P(y_j|x_j, \theta)\, P(x_j|\theta)$$

$$= \sum_{x_j} \{ \Phi^{X_j Y_j}_{x_j y_j} \} \exp\{ \sum_q (\theta_q - \delta_{jq}(x_j))B_{jq}(x_j)\}\, C(\theta,\delta_j)^{-1}$$

where $\delta_j = (\delta_{j1},....,\delta_{jv})$.

$$C(\theta,\delta_j) = \sum_z \exp \{ \sum_q (\theta_q - \delta_{jq}(z))B_{jq}(z)\}$$

As Kelderman (1988b) has shown, by specifying the scoring weights $B_{jq}(.)$, different models can be chosen for the dependence of the latent states on the latent traits. To illustrate the main idea of this paper, in the following one specific MPLT model will be considered; the multidimensional partial credit (MPCM) model. However, it may be clear that the contents of this paper is also valid for other kinds of MPLT models.

The scoring weights for a MPCM model, where each step depends on a different latent trait, are depicted in Figure 1(a).

Insert Figure 1 here

The "Complete knowledge" state (x=2) has scoring weight $B_{j1}(2) = 1$ on the first trait and scoring weight $B_{j2}(2) = 1$ on the second. The "Partial knowledge" state (x=1) has scoring weight $B_{j1}(1) = 1$ on the first trait, whereas the "Don't know" state (x=0) has scoring weights zero. Using (2) and the scoring weights $B_{jq}(.)$ of Figure 1(a), the probability that the subject has a latent state x_j on item j can be written as

$$(4) \qquad P(x_j|\theta) = \frac{\exp\{\sum_{q=1}^{x_j}(\theta_q - \delta_{jq}(x_j))\}}{\sum_z \exp\{\sum_{q=1}^{z}(\theta_q - \delta_{jq}(z))\}}$$

Adding a constant c to $\delta_{jq}(l)$ and subtracting it from $\delta_{jq}(l')$ $(1 \le l \ne l' \le x_j)$ does not change the model in (4). By setting the difficulty parameters of the same response equal to each other (i.e., $\delta_{jq}(x_j) = \delta_{jq}$ for $x_j = 0,1,2$ and all q), this indeterminancy can be removed (Kelderman, 1988b). From (4) and the assumption of local independence of the X_j's given the latent trait vector θ, it follows that the simultaneous distribution of X given θ is

$$(5) \qquad P(\mathbf{x}|\theta) = \exp\{\sum_q (\theta_q t_q - \sum_j B_{jq}(x_j)\,\delta_{jq})\} \prod_j C(\theta,\delta_j)^{-1}$$

where

$$C(\theta,\delta_j) = \sum_z \exp\{\sum_{q=1}^{z}(\theta_q - \delta_{jq})\}$$

and

$$t_q = \sum_j B_{jq}(x_j) \qquad q=1,\ldots,v$$

As all other MPLT models, the MPCM model is a exponential family model and $t=(t_1,\ldots,t_v)$ is a sufficient statistic for the latent trait vector θ (Kelderman, 1988b).

Let $\sum_x$ mean the summation over all possible latent state patterns $x=(x_1,\ldots,x_k)$. Using (1), (5) and the assumption of local independence of the observed responses y_j, the marginal probability of y given θ can be written as

(6) $$P(y|\theta) = \sum_x P(y|x,\theta)\, P(x|\theta)$$

$$= \sum_x \{ \prod_j \Phi_{x_j y_j}^{X_j Y_j} \} \exp\{ \sum_q (\theta_q t_q - \sum_j B_{jq}(x_j)\, \delta_{jq}) \} \prod_j C(\theta,\delta_j)^{-1}$$

Rearranging the terms of (4) and letting $F(\theta)$ be the distribution function of the latent ability vector θ, the marginal probability of the observed responses y can be written as an incomplete LCA model in the sense of Haberman (1979)

(7) $$P(y) = \sum_x \Phi_t^T \Phi_{x_1}^{X_1} \ldots \Phi_{x_k}^{X_k} \Phi_{x_1 y_1}^{X_1 Y_1} \ldots \Phi_{x_k y_k}^{X_k Y_k}$$

with

$$\Phi_t^T = \int \exp(\sum_q \theta_q t_q) \prod_j C(\theta,\delta_j)^{-1} dF(\theta),$$

$$\Phi_x^{X_j} = \exp(-\sum_{q=1}^{x} \delta_{jq}), \qquad (j=1,\ldots,k),$$

In this model, each value of the latent state vector **x** represents a latent class. Maximum likelihood estimates of the parameters of the GSERE model can be obtained by solving the maximum likelihood equations by the iterative proportional fitting (IPF) algorithm (Bartholomew, 1987; Goodman, 1974b; Haberman, 1979; Hagenaars, 1990). The overall goodness-of-fit of a model can be tested by the Pearson statistic or the likelihood-ratio statistic (see Haberman, 1979). Together with the question of identifiability, these two issues will be discussed in the next sections. But first, some applications of the GSERE model will be discussed.

Applications of the GSERE Model

In the previous section a GSERE model was formulated, where the parameters of the model were unrestricted, except for the usual restrictions pertaining to probabilities and conditional probabilities. In the present section, it is discussed how to modify the GSERE model in order to make it suitable for special applications.

Generally, for each specific GSERE model we may define the weights $B_{jq}(.)$ and certain constraints on the attraction parameters for each item j. The choices of the weights may depend on the required latent trait abilities for the correct response. For example, the item "20-5-6=?" requires two subtraction operations for the correct response, so that we can choose the one-

dimensional partial credit model as depicted in Figure 1(b). But, for the item "$\sqrt{(169-25)}=?$", where the two abilities are subtraction and taking the square root, we can choose the two-dimensional partial credit model as depicted in Figure 1(a). In Kelderman (1988b) other possible choices of the scoring weights for the dependence of the latent states on latent traits are discussed.

Insert Figure 2 here

By specifying the attraction parameters (1) as free, equal to each other or fixed to a certain value, a particular GSERE model is specified (Kelderman, 1988a; Westers & Kelderman, 1992). In Figure 2 some examples of the attraction parameters for the GSERE model are depicted. Figure 2(a) describes the situation of a perfect response process; the subject answers the item correctly (y=1) if (s)he can solve the problem (x=1) and gives a wrong answer (y_j=2) if (s)he cannot solve the problem (x=0). One case, where the items are not necessarily answered incorrectly if the subject cannot solve the problem, is with multiple choice items. If the subject doesn't know the answer, (s)he will guess the most attractive alternative. The attraction parameters for this situation are depicted in Figure 2(b). In Figure 2(c) the situation is depicted for the case where more then one alternative is correct. Generally, it is assumed that if the subject can solve the problem formulated by the item, (s)he will give the correct answer. But, a subject may fail to produce a correct answer, even if the subject was able to solve the problem. On the other hand, if (s)he is not able to solve the problem, it is impossible to produce the correct answer. Such a situation is depicted in Figure 2(d), where β is the so-called omission error. In the case of the MPCM model, we may assume that the attraction parameters are specified as depicted in Figure 2(e).

Identifiability

Whether the maximum likelihood estimates of the parameters of the GSERE model are unique depends on the identifiability of the model. A necessary condition for identifiability is, of course, that the number of independent parameters to be estimated does not exceed the number of cell frequencies minus one (i.e., (Πr_j)-1). Furthermore, if the MPLT model is not (locally) identifiable, then the GSERE model is also not (locally) identifiable.

Generally, in MPLT models identifying restrictions must be put on the model parameters. Therefore, in the paper of Kelderman (1988b) conditions are formulated, which ensures that the difficulty parameters of the MPLT model are not linearly dependent upon each other and upon the proportionality constants, respectively.

Since the (G)SERE model can be formulated as an (incomplete) LCA model, Goodmans (1974a) sufficient condition for identifiability can be used for the identifiability of the GSERE model. Let M be the matrix consisting of the derivatives of the function (7) with respect to the parameters of the GSERE model. The number of rows of the matrix M is equal to (Πr_j)-1 and the number of columns is equal to the number of parameters of the GSERE model. By direct extensions of a standard result about Jacobians, the GSERE model will be locally identifiable if the rank of the matrix M is equal to the number of columns. The rank of the matrix M can be evaluated by numerical methods.

When the parameters of the GSERE model are not identifiable, various kinds of restrictions can be imposed upon the attraction parameters and/or the item parameters in order to make the GSERE model identifiable. For instance, by specifying the attraction parameters to be equal to each other or to be fixed at a certain value.

Unidentifiability can be discovered by estimating the parameters a second time, this time using different initial estimates. In the case of unidentifiability both runs will give different parameter estimates.

Estimation Method

Let m_{xyt} be the expected number of subjects with latent state x, observed response y and sumscore t. As Kelderman (1988a), and Westers and Kelderman (1992) have showed, the parameters of the SERE model can be estimated by applying the iterative proportional fitting (IPF) algorithm. For the GSERE the IPF algorithm can also be used. One of the differences with the SERE model is that, depending on the number of the latent state categories, the number of latent classes may be quite large. Since the number of attraction parameters depends on the number of latent state categories and the number of item response categories, this number may also be quite large.

The maximum likelihood estimates of the parameters of the GSERE model can be obtained by solving the maximum likelihood equations by a two-step algorithm. In the first step of each iteration (i.e., the outer iteration), the attractiveness of the alternatives and the expected frequency distribution of the latent classes will be estimated. For the GSERE model the first step is similar to the first step of the estimation method for the parameters in the LCA model (Goodman, 1974b). In the second step of each iteration (i.e., the inner iteration), the estimated expected distribution of the latent classes is fitted to the postulated loglinear model. From this distribution the difficulty parameters can be estimated.

5.1 Outer Iteration

As indicated before the GSERE model can be formulated as a LCA model, where each latent class represents a latent state vector x. Let

$$(8) \qquad P(y) = \sum_{x} P(x)\, P(y_1|x) \ldots P(y_k|x)$$

where $P(\mathbf{x})$ is the probability of getting latent state vector $\mathbf{x}$ and $P(y_j|\mathbf{x})$ is the conditional probability of choosing response y_j given the latent state vector $\mathbf{x}$. The model in (8) is a LCA model in the sense of Goodman (1974b), which means that the IPF algorithm for the general latent structure model, where the parameters in the model are unrestricted, can be used for estimating the expected frequency distribution of the latent classes and the attractiveness of the alternatives.

As mentioned before, in the GSERE model it is assumed that the observed response y_j depends only on the latent state x_j, therefore the conditional probabilities $P(y_j|\mathbf{x})$ are restricted in the following manner

$$P(y_j|\mathbf{x}) = P(y_j|\mathbf{x}') = \Phi^{X_j Y_j}_{x_j y_j}$$

for all latent state vectors $\mathbf{x}$ and $\mathbf{x}'$ with components $x'_j = x_j$. The Φ^{XY} can be obtained by taking a weighted average of the estimates $\hat{P}(y_j|\mathbf{x})$ obtained from the IPF algorithm, using weights proportional to $\hat{P}(\mathbf{x})$ (Goodman, 1974b). So

$$\Phi^{X_j Y_j}_{x_j y_j} = \{ \sum_{\mathbf{x}.} \hat{p}(\mathbf{x}) \hat{p}(y_j|\mathbf{x}) \} / \{ \sum_{\mathbf{x}.} \hat{p}(\mathbf{x}) \} ,$$

where $\sum_{\mathbf{x}.}$ means the summation over all latent state vectors $\mathbf{X}=\mathbf{x}$ with $X_j=x_j$.

5.2 Inner Iteration

As assumed in the second section of this paper, the latent probabilities $P(\mathbf{x})$ are restricted in such a way that they satisfy an MPCM model. Knowing that the MPCM model is an exponential family model and that the sum score $\mathbf{t}$ is a sufficient statistic for the latent ability parameter θ (Kelderman, 1988b), the conditional distribution of $\mathbf{X}$ given $\mathbf{t}$ is by

$$P(\mathbf{x}|\mathbf{t}) = \exp\{ \sum_{j} \varphi_j(x_j) \} / g(\mathbf{t},\varphi),$$

with

$$g(\mathbf{t},\varphi) = \sum_{\mathbf{x}|\mathbf{t}} \exp\{ \sum \varphi_j(x_j) \},$$

where $\sum_{\mathbf{x}|\mathbf{t}}$ means summation over those values of the latent state vector $\mathbf{x}$ for which $(\sum B_{j1}(\mathbf{x}),....,\sum B_{jv}(\mathbf{x}))$ is equal to $\mathbf{t}$, and the vector $\varphi = (\varphi_1(x_1),....,\varphi_k(x_k))$ denotes the weighted sums over latent traits of the difficulty parameters

$$\varphi_j(x_j) = - \sum_{q=1}^{v} \delta_{jq}(x_j)\, B_{jq}(x_j).$$

Letting $m_{\mathbf{xt}}$ be the estimated expected number of subjects with latent state $\mathbf{x}$ and sum score $\mathbf{t}$ we have

(9) $$\log m_{\mathbf{xt}} = \sigma_{\mathbf{t}} + \sum_{j} \varphi_j(x_j)$$

where $\sigma_{\mathbf{t}} = - \log(g(\mathbf{t},\varphi)/N_{\mathbf{t}})$ is a fixed normalizing constant, $N_{\mathbf{t}}$ is the number of subjects with sum score $\mathbf{t}$. For the MPCM model the parameters $\varphi_j(x_j)$ are specified by $\varphi_j(1) = \delta_{j1}$ and $\varphi_j(2) = -\delta_{j1} - \delta_{j2}$, respectively.

The model in (9) is a quasi-loglinear model for an incomplete item response 1 x...x item response k x score 1 x...x score v contingency table. The table is incomplete since for certain given values of $\mathbf{X}$, only one $\mathbf{t}$ is possible. Maximum likelihood estimates can be obtained by solving the maximum likelihood equations of the MPCM model. These equations can be solved by IPF (Kelderman, 1988b). The latent probabilities $P(\mathbf{x})$ are then adjusted to these maximum

likelihood estimates. In this way new latent probabilities P(x) are obtained that satisfied the postulated MPLT model and are used again in the next outer iteration.

Goodness-of-Fit Test

The overall goodness of fit of the GSERE model can be tested by the Pearson statistic or the likelihood-ratio statistic. Both statistics are asymptotically distributed as chi square with degrees of freedom equal to the difference between the number of cells in the observed contingency table and the number of estimable parameters. However, if the expected counts become too small, the approximation of the distribution of the goodness-of-fit statistics by a chi-square distribution becomes bad (Koehler, 1977; Lancaster, 1961; Lamz, 1978).

Using the difference in the likelihood-ratio test statistics for two models (Bishop, Fienberg, & Holland, 1975; Rao, 1973), it can be tested whether an alternative model yields a significant improvement in fit over a special case of this alternative model.

If two GSERE models are not proper subsets of each other, then Akaike's (1977) information criterion (AIC) or Raftery's (1986a, 1986b) Bayesian information criterion (BIC) can be used. AIC is defined as

$$AIC = G^2 - (\ln 2)\, d$$

where G^2 is the likelihood-ratio test statistic and d is the number of independent parameters in the GSERE model. The BIC index has ln N (i.e., the sample size) instead of ln 2, but is otherwise identical. For both indices, the first term is a measure of badness of fit, whereas the second term is a correction for overfitting due to the increasing bias in G^2 as the number of parameters in the SERE model increases. The GSERE with the minimum AIC or BIC value will be chosen to be

the best fitting model. Computer programs by Hagenaars and Luijkx (LCAG, 1990) and Kelderman and Steen (LOGIMO, 1988) can be used to fit the model.

Example

For numerical reasons, the GSERE model is difficult to be routinely applied in large testing programs yet. Not the number of parameters of the model causes the problems, but the number of latent classes and the tables of observed and expected counts become too large for computer storage. One solution to this problem may be the use of the Division-by-Item (DBI) principle of Westers (1992). In this paper a maximum likelihood estimation method for the one-dimensional SERE model for a large set of items was described. This method was based on dividing the whole item set in several subsets. The computational problem boils down to the simultaneous estimation of the parameters of a set of smaller SERE models. This could be done because one of the properties of the SERE model was the collapsibility of the model. Since this property is also valid for the GSERE model (Appendix A.2), we can also use the DBI-principle for the estimation of the parameters of the GSERE model. However, before the GSERE model can be widely applied, further research is required to reduce the amount of computer storage. But if the generalization to the multidimensional latent spare is ignored, then, for a small number of items, the parameters of the GSERE model can still be estimated by the combined of the programs LCAC and LOGIMO. In this section this will be demonstrated with an example.

The authors were allowed to analyse four one-dimensional four-choice items from a protected data base of the ASCP Medical Laboratory Test. ASCP produces tests for certification of paramedical personnel. The items in the Medical Laboratory Test measures the ability to perform medical laboratory tests. The test score is obtained by adding the number of correct items.

Content experts from ASCP have hypothesized that the cognitive process "Applying Knowledge" was involved in answering these four items. According to the assumptions of the

---

Insert Table 1 here

---

content experts, we postulated that there are three latent states: "Don't know", "Partial knowledge" and "Complete knowledge", with scoring weights equal to zero, one and two, respectively. This means that it is assumed that the latent states are governed by the one-dimensional partial credit model (OPCM) as depicted in Figure 1(b). Table 1 shows the hypothesized weights that content experts gave for each of the item responses on the cognitive

---

Insert Figure 3 here

---

process "Applying Knowledge". Those hypothesized scoring weights are translated into specifications of the attraction parameters, which are depicted in Figure 3(a) through (d), where x=0 indicates the "Don't know" state, x=1 the "Partial knowledge" state and x=3 the "Complete knowledge" state, respectively.

In this example the hypothesized model (H) will be compared which three alternative models. The first model (A_1) is a GSERE model, where for each item all the attraction parameters are unequal to zero (Figure 4). Model A_2 is the same as Model A_1, but where it is assumed that the correct alternative will be chosen if the subject is in the "Complete knowledge" state. Model A_3 has not only the same assumptions as Model A_2, but it also assumes that $\alpha_i=\beta_i$ (i=1,...,4). The specifications of the attraction parameters of Models A_2 and A_3 are also depicted in Figure 4.

Insert Figure 4 here

In Table 2 the Likelihood-ratio test statistics, the Akaike's information criteria and Raftery's Bayesian information criteria for the four models are given. From the G^2 and AIC values it is seen that the hypothesized model fits the data better then the alternative models. Furthermore, Model A_2 fits the data better than Model A_3, which means that there may be a significant difference between the attraction of the alternatives for a subject in the "Don't know" state and a subject in the "Partial knowledge" state. The better fit of Model A_2 relative to Model A_1 may indicate that a subject in the "Complete knowledge" state would make no mistake choosing the right alternative.

Insert Table 2 here

In Table 3 the estimates of the attraction parameters for the alternatives of each item are presented for the hypothesized model. These results indicate that a subject in the "Partial knowledge" state is more likely to choose the correct alternatives to Items 1 and 4 than a subject in the "Don't know" state. Most likely, a test constructor would never expect that a subject in the "Don't know" state is more likely to choose the correct alternative than a subject in the "Partial knowledge" state, as has happened for Item 2. The attraction parameter of the correct alternative of Item 2 in the first state is five times as large as the associated parameter in the second state. However, in both states the probability of choosing the right alternative is very low. In the "Partial Knowledge" state alternative A is almost always chosen, whereas in the "Don't know"

state alternative D is often chosen. Therefore, Item 2 has the advantage of being possible to distinguish between subjects who exactly knew the solution of the problem imposed by Item 2 and those who did not. The attraction parameters for the correct alternatives of Item 3 are approximately the same for both states.

From Table 3 it can be seen that some of the attraction parameters are smaller than .05. An interesting question for these cases is if these attraction parameters are really unequal to zero or happen to yield estimates unequal to zero by chance. Although it is bad practice to formulate an alternative model post hoc after looking at parameter estimates and test it on the sample sample, we have tried to find an answer to the above mentioned question by fitting a fourth alternative model (A_4), which had the same assumptions as the hypothesized model (H). It was also assumed that all the attraction parameters with estimated value smaller than .05 in the hypothesized model were equal to zero. This alternative model showed a slightly improved fit

---

Insert Table 3 here

---

compared to the hypothesized model (see Table 2). It also followed that the alternative model fits the data better than all the other alternative models. In Table 4 the estimated attraction parameters for Model A_4 are given. From this table it can be seen that the estimated attraction parameters in the case of Model A_4 do not differ much from the estimated attraction parameters of the hypothesized model.

Insert Table 4 here

At this point we have considered only the attraction parameters of the four items from the ASCP Medical Laboratory Test. In the remaining part of this section we will take a closer look at the difficulty parameters of the items. In this example it was assumed that the latent states were governed by the one-dimensional partial credit model. Using the scoring weights for this model as depicted in Figure 1(b) and omitting the latent trait index q, the one-dimensional version of Model (2) becomes

$$(10) \qquad P(x|\theta) = \frac{\exp((\theta - \delta_j(x))x)}{\sum_z \exp((\theta - \delta_j(z))z)}$$

$$= \frac{\exp(\sum_{g=1}^{x} (\theta - \psi_{jg}))}{\sum_z \exp(\sum_{g=1}^{z} (\theta - \psi_{jg}))}$$

where $\psi_{jx} = x\,\delta_j(x) - (x-1)\,\delta_j(x-1)$ describes the difficulty of step x in item j, because each latent state may be seen as the result of a series of subsequent steps, each of which has to be passed. In Table 5 the values of the ψ_{jx} parameters for the four items of the ASCP Medical Laboratory Test in the case of the hypothesized model H are given.

Insert Table 5 here

From Table 5 it follows that the difficulty of the steps changes positively for the items 1, 3 and 4. Meaning that it is more difficult to do the last step than the first step. On the other hand, for item 2 it is difficult to take the first step, but if the first step is reached than the second step is very

easy. Following Verhelst and Verstralen (1991), two remarks have to be made. First, as Molenaar (1983) showed, the parameter value of a particular step depends on the parameter values for the other steps in the item. Therefore, the parameter value of a step cannot be interpreted as a measure of its difficulty. Second, it cannot be known in advance that the items allow for a sequential solution as assumed in the partial credit model.

Finally, some surprising results which are found during the analyses will be discussed. First from the 3370 subjects in the study no one had responded complete wrong to the four items. However, in the case of the hypothesized model it was estimated that 68 subjects had no knowledge of the solution of the problems imposed by all four items. On the other hand, almost 32% of the subjects responded complete right to all four items, whereas it was estimated that 2% really knew all the solutions to the problems.

Discussion

In this paper a loglinear item response theory (IRT) model with latent classes was proposed that related polytomously scored item responses to a multidimensional latent space. The proposed model is a generalization of the solution-error response-error (SERE) model (Kelderman 1988a; Westers & Kelderman, 1992) to situations of polytomously scored latent states that may be explained by a multidimensional latent space. In this generalized SERE model a distinction is made between some well-defined latent states in which the subject has a certain amount of knowledge of the answer. The probability that the subject is in a certain state is assumed to be governed by the multidimensional polytomous latent trait model (MPLT). The relationship between the latent states and the observed answers is described by conditional probabilities.

Maximum likelihood estimates of the parameters of the GSERE model can be obtained by the IPF algorithm. However, the results by Westers and Kelderman (1992) indicate that the (M)SERE model is usable in practice only when the responses to a few items are studied.

However, since the property of collapsibility is also valid for the GSERE model, the DBI-principle of Westers (1992) can be used for estimating larger sets of parameters in the GSERE model.

As pointed out in Westers and Kelderman (1992), an item can show DIF in two different ways. First, an item exhibits DIF if equally able individuals from different subgroups have different probabilities of knowing the answer. Second, an item also exhibits DIF if the attractiveness of the alternatives vary from subgroup to subgroup. Just as in the case of the SERE model, the GSERE model can be extended with variables defining subgroups to study these two types of DIF. Therefore, the GSERE model is suitable for the examination of DIF in polytomous items through a combination of DIF for correct/incorrect responses and DIF in the alternatives.

Appendix A.1

Two Items of the American Society of Clinical Pathologist (ASCP) Microbiology Test

Item 11 Of the following bacteria, the most frequent cause of prosthetic heart valve infections occurring two to three months after surgery is:

a. Streptococcus pneumoniae
b. Streptococcus pyogenes
c. Staphylococcus aureus
d. Staphylococcus epidermidis

Item 20 A beta-hemolytic gram-positiv coccus was isolated from the cerebrospinal fluid of a 2-day-old infant with signs of meningtis. The isoltae grew on sheep blood agar under aerobic conditions and was resistant to a bacitracin disc. Which of the following should be performed for the presumptive identification of the organism?

a. oxidase production
b. catalase formation
c. CAMP test
d. esculin hydrolysis

Appendix A.2

The Collapsed Generalized Solution-Error

Response-Error Model

In this appendix an instructional proof will be given that the GMSERE model in (7) is collapsible. The same notation will be used as in the previous sections. From elementary probability theory, it follows for the GSERE model that

$$P(y_1,y_2) = \sum_{y_3} \dots \sum_{y_k} P(y_1,\dots,y_k) \equiv \sum_{y_3} \dots \sum_{y_k} P(\mathbf{y})$$

$$= \sum_{\mathbf{x}} \sum_{y_3} \dots \sum_{y_k} P(\mathbf{x})\, P(\mathbf{y}|\mathbf{x}) = \sum_{\mathbf{x}} P(\mathbf{x}) \sum_{y_3} \dots \sum_{y_k} P(\mathbf{y}|\mathbf{x})$$

$$= \sum_{\mathbf{x}} P(\mathbf{x})\, P(y_1|x_1)\, P(y_2|x_2) \sum_{y_3} \dots \sum_{y_k} P(y_3|x_3) \dots P(y_k|x_k)$$

with $P(\mathbf{x})$ is the marginal probability of the latent state vector $\mathbf{x}$. From conditional probability calculus it follows that

$$P(y_1,y_2) = \sum_{\mathbf{x}} P(\mathbf{x})\; P(y_1|x_1)\, P(y_2|x_2)$$

$$= \sum_{x_1} \sum_{x_2} P(y_1|x_1)\, P(y_2|x_2) \sum_{x_3} \dots \sum_{x_k} P(\mathbf{x})$$

Letting $\mathbf{z}=(z_1,\dots,z_v)$ with $z_q = B_{1q}(x_1) + B_{2q}(x_2)$, and using the assumption of local independence of the latent states x_j in the MPCM model and elementary calculus it follows that

$$\sum_{x_3} \cdots \sum_{x_k} P(\mathbf{x}) = \sum_{x_3} \cdots \sum_{x_k} \int P(\mathbf{x}|\theta)\, dF(\theta)$$

$$= \int \sum_{x_3} \cdots \sum_{x_k} \prod_{j=1}^{k} P(x_j|\theta)\, dF(\theta)$$

$$= \int P(x_1|\theta)\, P(x_2|\theta)\, dF(\theta)$$

Using this last equation, the marginal probability of the observed responses y_1 and y_2 can be written as

$$P(y_1,y_2) = \sum_{x_1} \sum_{x_2} P(y_1|x_1)\, P(y_2|x_2) \int P(x_1|\theta)\, P(x_2|\theta)\, dF(\theta)$$

$$= \sum_{x_1} \sum_{x_2} \Phi^{X_1 Y_1}_{x_1 y_1} \Phi^{X_2 Y_2}_{x_2 y_2} \Phi^{X_1}_{x_1} \Phi^{X_2}_{x_2} \Phi^{T_{12}}_{z}$$

with

$$\Phi^{T_{12}}_{z} = \int \exp\left(\sum_{q} z_q \theta_q\right) \{C(\theta,\delta_1)\, C(\theta,\delta_2)\}^{-1}\, dF(\theta)$$

and

$$C(\theta,\delta_j) = \sum_{z} \exp\left\{ \sum_{q=1}^{z} (\theta_q - \delta_{jq}) \right\}$$

which is similar to (7) except that here we consider two items and in (7) k items. It may be clear that the collapsibility of the GSERE model is also valid, if the general MPLT model is used for describing the dependence of the latent states on the latent traits.

References

Akaike, H. (1977). On entropy maximizations principle. In P.R. Krisschnaiah, ***Application of statistics*** (p. 27-41). North Holland.

Bartholomew, D.J. (1987). ***Latent variable models and factor analysis***. London: Griffin.

Bishop, Y.M.M., Fienberg, S.E., & Holland, P.W. (1975). ***Discrete multivariate analysis***. Cambridge, MA: MIT Press.

Bock, R.D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. ***Psychometrika***, ***46***, 443-459.

Cressie, N., & Holland, P.W. (1983). Characterising the manifest probabilities of latent trait models. ***Psychometrika***, ***48***, 129-142.

Dayton, C.M., & Macready, G.B. (1980). A scaling model with response errors and intrinsically unscalable respondents. ***Psychometrika***, ***45***, 343-356.

Duncan, O.D. (1984). Rasch measurement: Further examples and discussion. In C.F. Turner, & E. Martin (Eds.), ***Surveying subjective phenomena, Vol. 2*** (p. 367-403). New York: Russell Sage Foundation.

Formann, A.K. (1985). Constrained latent class models: Theory and applications. ***British Journal of Mathematical and Statistical Psychology***, ***38***, 87-111.

Goodman, L.A. (1974a). Exploratory latent structure analysis using both identifiable and unidentifiable models. ***Biometrika***, ***61***, 215-231.

Goodman, L.A. (1974b). The analysis of systems of qualitative variables when some of the variables are unobservable. Part I - a modified latent structure approach. ***The American Journal of Sociology***, ***79***, 1179-1259.

Haberman, S.J. (1979). ***Analysis of qualitative data. Vol. 2. New Developments***. New York: Academic Press.

Hagenaars, J. (1990). ***Categorical longitudinal data: Log-linear panel, trend, and cohort analysis***. Newbury Park, CA: Sage Publications.

Hagenaars, J., & Luijkx, R. (1990). ***LCAG: latent class models and other loglinear models with latent variables***. (Working paper # 17^2). Tilburg: Tilburg University.

Kelderman, H. (1984). Loglinear Rasch model tests. ***Psychometrika, 49***, 223-245.

Kelderman, H. (1988a). ***An IRT model for item responses that are subject to omission and/or intrusion Errors***. (Research Report 88-16). Enschede: University of Twente.

Kelderman, H. (1988b). ***Loglinear multidimensional IRT model for polytomously scored items***. (Research Report 88-17). Enschede: University of Twente.

Kelderman, H. (1989). Item bias detection using loglinear IRT. ***Psychometrika, 54***, 681-697.

Kelderman, H., & Macready, G.B. (1990). The use of loglinear models for assessing differential item functioning across manifest and latent examinee groups. ***Journal of Educational Measurement, 27***, 307-327.

Kelderman, H., & Steen, R. (1988). ***LOGIMO I: Loglinear item response theory modeling***. [Computer program]. Enschede: University of Twente, Department of Educational Technology.

Koehler, K.J. (1977). ***Goodness-of-fit statistics for large sparse multinomials***. Doctoral dissertation. School of Statistics, University of Minnesota.

Lancaster, H.O. (1961). Significance tests in discrete distributions. ***Journal of the American Statistical Association, 56***, 223-234.

Larnz, K. (1978). Small sample comparisons of exact levels for chi-square goodness-of-fit statistics. ***Journal of the American Statistical Association, 73***, 252-263.

Lazarsfeld, P.F., & Henry, N.W. (1968). ***Latent structure analysis***. Boston: Houghton-Miffin.

Lord, F.M. (1980). ***Applications of item response theory to practical testing problems***. Hillsdale, NJ: Lawrence Erlbaum.

Lord, F.M., & Novick, M.R. (1968). ***Statistical theories of mental test scores***. Reading, MA: Addison-Wesley.

Mislevy, R.J., & Verhelst, N. (1990). Modeling item responses when different subjects employ different solutions strategies. ***Psychometrika, 55***, 195-216.

Molenaar, I.W. (1983). Some improved diagnostics for the Rasch model. ***Psychometrika***, ***48***, 49-72.

Mooijaart, A. (1978). ***Latent structure models***. Doctoral dissertation. Leiden: University of Leiden.

Raftery, A.E. (1986a). Choosing models for cross-classifications. ***American Sociological Review***, ***51***, 145-146.

Raftery, A.E. (1986b). A note on Bayes factors for log-linear contingency table methods with vague prior information. ***Journal of the Royal Statistical Society. Series B***, ***48***, 249-250.

Rao, C.R. (1973). ***Linear statistical inference and its applications***. New York: Wiley.

Rasch, G. (1960/1980). ***Probabilistic models for some intelligence and attainment tests***. Copenhagen: Paedagogiske Institut. (Expanded edition, Chicago: The University of Chicago Press, 1980).

Tjur, T. (1982). A connection between Rasch's item analysis model and a multiplicative Poisson model. ***Scandinavian Journal of Statistics***, ***9***, 23-30.

Verhelst, N., & Verstralen, H.H.F.M. (1991). ***The partial credit model with non-sequential solution strategies***. (Measurement and Research Department Reports 91-5). Arnhem: Cito

Westers, P. (1992). ***Estimating the parameters of the solution-error response-error model using subsets of items***. Unpublished manuscript. Enschede: University of Twente.

Westers, P., & Kelderman, H. (1992). Examining differential item functioning due to item difficulties and alternative attractiveness. ***Psychometrika***, ***57***, 107-118.

Yamamoto, K. (1987). ***A model that combines IRT and latent class models***. Illinois: University of Illinois at Urbana-Champaign.

Yamamoto, K. (1988). ***Hybrid model of IRT and latent class models***. Princeton, NJ: Educational Testing Service.

Table 1

Hypothesized Weights of the ASCP Medical Laboratory Test Items for the Cognitive Process "Applying Knowledge".

Item	Scoring Weights				Correct Answer
	A	B	C	D	
1			2	1	C
2	1		2		C
3	2	1	1		A
4	2	1	1		A

Table 2

Likelihood-Ratio Test Statistics, Akaike's Information Criteria and Raftery's Bayesian Information Criteria for Four Items from the ASCP Medical Laboratory Test

Model	Number of Parameters	G^2 Value	AIC Value	BIC Value
H	35	197.557	173.297	-86.736
A_1	53	328.733	291.996	-101.768
A_2	41	190.259	161.840	-142.770
A_3	29	235.102	215.001	-0.455
A_4	32	200.283	178.102	-59.642

Table 3

Attraction Parameters for the Alternatives of Four Items from the ASCP Medical Laboratory Test in the case of Model H

	Alternatives "Don't know" state				Alternatives "Partial knowledge" state			
Item	A	B	C	D	A	B	C	D
1	.204	.135	.180	.481	.000*	.000*	.676	.324
2	.027	.276	.064	.633	.988	.000*	.012	.000*
3	.622	.163	.000	.215	.502	.158	.340	.000*
4	.556	.120	.086	.238	.828	.110	.063	.000*

Note 1.: The correct alternatives are underlined.

Note 2.: Attraction parameters marked with an asterisk are prespecified.

Table 4

Attraction Parameters for the Alternatives of Four Items from the ASCP Medical Laboratory Test in the case of Model A_4

	Alternatives "Don't know" state				Alternatives "Partial knowledge" state			
Item	A	B	C	D	A	B	C	D
1	.275	.182	.187	.355	.000*	.000*	.581	.419
2	.000*	.286	.058	.656	1.000	.000*	.000	.000*
3	.623	.164	.000*	.208	.597	.124	.279	.000*
4	.477	.130	.112	.281	.919	.081	.000	.000*

Note 1.: The correct alternatives are underlined.

Note 2.: Attraction parameters marked with an asterisk are prespecified.

Table 5

Estimated Difficulty Parameters of Four Items from the ASCP Medical Laboratory Test for the case of the Hypothesized Model H

Item	Step 1	Step 2
1	-2.15232	-0.64080
2	0.18140	-4.02920
3	-1.87987	-0.51876
4	-2.04992	1.20943

Figure Captions

Figure 1. Scoring Weights for the One- and Two-dimensional Partial Credit Model.

Figure 2. Examples of Specifications of the Attraction Parameters

Figure 3. Specifications of the Attraction Parameters of the Four Items from the ASCP Medical Laboratory Test for the Hypothesized Model H.

Figure 4. Specifications of the Attraction Parameters of the Four Items from the ASCP Medical Laboratory Test for the Alternative Models A_1, A_2, and A_3.

Note 1. y=A expresses the right alternative.

Note 1. The specifications are for all four items the same.

		q	
		1	2
	0	0	0
x	1	1	0
	2	1	1

(a)

		q
		1
	0	0
x	1	1
	2	2

(b)

x \ y	1	2
0	0	1
1	1	0

(a)

x \ y	1	2	3	4
0	α_1	α_2	α_3	α_4
1	1	0	0	0

(b)

x \ y	1	2	3	4
0	α_1	α_2	α_3	α_4
1	β_1	0	β_2	0

(c)

x \ y	1	2
0	0	1
1	1-β	β

(d)

x \ y	1	2	3	4
0	α_1	α_2	α_3	α_4
1	β_1	0	β_2	0
2	1	0	0	0

(e)

x \ y	A	B	C	D
0	α_1	α_2	α_3	α_4
1	0	0	β_1	β_2
2	0	0	1	0

(a)

x \ y	A	B	C	D
0	α_1	α_2	α_3	α_4
1	β_1	0	β_2	0
2	0	0	1	0

(b)

x \ y	A	B	C	D
0	α_1	α_2	α_3	α_4
1	β_1	β_2	β_3	0
2	1	0	0	0

(c)

x \ y	A	B	C	D
0	α_1	α_2	α_3	α_4
1	β_1	β_2	β_3	0
2	1	0	0	0

(d)

x \ y	A	B	C	D
0	α_1	α_2	α_3	α_4
1	β_1	β_2	β_3	β_4
2	γ_1	γ_2	γ_3	γ_4

(A_1)

x \ y	A	B	C	D
0	α_1	α_2	α_3	α_4
1	β_1	β_2	β_3	β_4
2	1	0	0	0

(A_2)

x \ y	A	B	C	D
0	α_1	α_2	α_3	α_4
1	α_1	α_2	α_3	α_4
2	1	0	0	0

(A_3)

Titles of recent Research Reports from the Division of
Educational Measurement and Data Analysis,
University of Twente, Enschede,
The Netherlands.

RR-93-1 P. Westers & H. Kelderman, *Generalizations of the Solution-Error Response-Error Model*

RR-91-1 H. Kelderman, *Computing Maximum Likelihood Estimates of Loglinear Models from Marginal Sums with Special Attention to Loglinear Item Response Theory*

RR-90-8 M.P.F. Berger & D.L. Knol, *On the Assessment of Dimensionality in Multidimensional Item Response Theory Models*

RR-90-7 E. Boekkooi-Timminga, *A Method for Designing IRT-based Item Banks*

RR-90-6 J.J. Adema, *The Construction of Weakly Parallel Tests by Mathematical Programming*

RR-90-5 J.J. Adema, *A Revised Simplex Method for Test Construction Problems*

RR-90-4 J.J. Adema, *Methods and Models for the Construction of Weakly Parallel Tests*

RR-90-2 H. Tobi, *Item Response Theory at subject- and group-level*

RR-90-1 P. Westers & H. Kelderman, *Differential item functioning in multiple choice items*

Research Reports can be obtained at costs from Bibliotheek, Department of Education, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.

department of
EDUCATION

A publication by
the Department of Education
of the University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

BEST COPY AVAILABLE